

**The Report Committee for Felipe Ferreira Dias certifies
that this is the approved version of the following report:**

**A Behavioral Choice Model Of The Use Of
Car-Sharing And Ride-Sourcing Services**

Committee:

Tse-Min Lin, Supervisor

Chandra R. Bhat

**A Behavioral Choice Model Of The Use Of
Car-Sharing And Ride-Sourcing Services**

by

Felipe Ferreira Dias, B.Sc., M.Sc.

Report

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

**Master of Science in Statistics
The University of Texas at Austin
May, 2017**

Acknowledgements

Firstly, I would like to acknowledge the people who helped me co-author this research: Patrícia Lavieri, Dr. Venu M. Garikapati, Sebastian Astroza, Dr. Ram M. Pendyala and Dr. Chandra R. Bhat.

This research was partially supported by the U.S. Department of Transportation through the Data-Supported Transportation Operations and Planning (D-STOP) Tier 1 University Transportation Center. Dr. Bhat would also like to acknowledge support from a Humboldt Research Award from the Alexander von Humboldt Foundation, Germany.

I am also very grateful to Mr. Neil Kilgren of the Puget Sound Regional Council for providing the data used in this research study, and to Lisa Macias for all her help and support.

I would like to acknowledge the reviewers, whose comments added great depth and significance to this research project.

I am also deeply grateful to Dr. Tse-Min Lin, who accepted to be my supervisor and guided me through the final steps of my Master's degree.

A Behavioral Choice Model Of The Use Of Car-Sharing And Ride-Sourcing Services

Felipe Ferreira Dias, MSStat

The University of Texas at Austin, 2017

Supervisor: Tse-Min Lin

Abstract: There are a number of disruptive mobility services that are increasingly finding their way into the marketplace. Two key examples of such services are car-sharing services and ride-sourcing services. In an effort to better understand the influence of various exogenous socio-economic and demographic variables on the frequency of use of ride-sourcing and car-sharing services, this paper presents a bivariate ordered probit model estimated on a survey data set derived from the 2014-2015 Puget Sound Regional Travel Study. Model estimation results show that users of these services tend to be young, well-educated, higher-income, working individuals residing in higher-density areas. There are significant interaction effects reflecting the influence of children and the built environment on disruptive mobility service usage. The model developed in this paper provides key insights into factors affecting market penetration of these services, and can be integrated in larger travel forecasting model systems to better predict the adoption and use of mobility-on-demand services.

Table of Contents

List of Tables	vi
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 MOBILITY-AS-A-SERVICE AND TRAVEL DEMAND	4
CHAPTER 3 DATA DESCRIPTION	8
CHAPTER 4 MODELLING METHODOLOGY	11
CHAPTER 5 MODEL ESTIMATION RESULTS	14
CHAPTER 6 COMPUTATION OF PSEUDO-ELASTICITIES	19
CHAPTER 7 CONCLUSIONS	23
APPENDIX A: GAUSS CODE FOR MODEL ESTIMATION	25
APPENDIX B: PYTHON CODE FOR CALCULATION OF PSEUDO-ELASITCITIES	33
REFERENCES.....	46

List of Tables

Table 1 Survey Sample Description	10
Table 2 Estimation Results for Bivariate Ordered Probit Model	18
Table 3 Computation of Pseudo-Elasticity Effects	22

CHAPTER 1 INTRODUCTION

New mobility-on-demand services are transforming the transportation ecosystem. Two key developments in this arena include ride-sourcing services and most recent versions of car-sharing services. Ride-sourcing services, also referred to as transportation network companies (TNCs), real-time ride-sharing, parataxis, ride-hailing, and on-demand rides (Rayle et al., 2016) came into existence within the past decade and integrate a wide array of technological capabilities in a single package, offering users a mode of transportation that is analogous to the taxi, but at a lower cost. Ride-sourcing services offer reliable, lower cost (than traditional taxi services), on-demand, and door-to-door transportation that is requested (hailed), tracked, and paid by users through smartphone apps. Two key examples of these services are Uber and Lyft, although there are a number of country-specific examples that are growing at a rapid pace (e.g., Didi in China and Ola in India). Uber is the largest and most well-known mobility-on-demand ride-sourcing service provider, with presence in 450 cities around the world (Somerville, 2016). The provider served its two-billionth ride in July 2016, just six months after serving its one-billionth ride. Given that it took Uber six years to reach the one billion trip milestone, but only six months after that to reach the two billion trip mark, it is fair to say that ride-sourcing services are gaining popularity at a rapid pace around the world with strong growth both in drivers and riders. One poll shows that about 12 percent of registered voters across the United States use ride-sourcing services at least once a month (Morning Consult, 2015),

This work was originally presented at the Transportation Research Board's Annual Meeting of 2017. It was then submitted to the journal "Transportation" on March 21st, 2017 and is expected to be published in one of the journal's special editions.

The original paper was the result of a collective effort between several authors. Felipe Dias's main contributions to this work were structuring the paper as a whole, compiling the main literature review findings, performing some of the data cleaning and processing, estimating the model, writing the initial explanation of results and developing the Python code for the pseudo-elasticities. Sebastian Astroza and Patricia Lavieri helped with the literature review and the data cleaning and processing portions, while Dr. Garikapati and Dr. Pendyala expanded both the literature review and the explanation of the results. Dr. Bhat wrote the original GAUSS code that allowed for the model estimation and the methodology behind the pseudo-elasticities, while guiding the interpretation of the pseudo-elasticities' results as well as the project's conclusions.

while another survey showed that 25 percent of San Francisco residents used TNC services on a monthly basis (SFMTA, 2014).

The second major mobility service that is of interest within the scope of this paper is that of modern car-sharing services, which have also moved into an era of smartphone-based apps. These services allow users to benefit from the use of an automobile while avoiding the burden of private vehicle ownership (Shaheen et al., 2009). Two key examples of car-sharing services include ZipCar and car2go, among others. ZipCar is a car-sharing service which has dedicated parking spots at strategic locations throughout a city; users are expected to pick-up and drop-off (park) the shared cars at these specific parking locations. On the other hand, car2go is a free-floating system in which users may pick-up and drop-off cars wherever they please as long as they are not violating local parking and traffic ordinances. Although these services have not grown at the same pace as ride-sourcing services, they are gaining in popularity and a number of car-sharing services are present in cities around the world. In the Americas alone, there were an estimated 22,000 shared vehicles reaching 1.5 million car-sharing members in 2015 (Shaheen, 2016). The role, use, and impact of car-sharing services have been studied more than that of ride-sourcing services; however, there is still much to be learned about the adoption and use of these services, particularly in the context of their new smartphone-based incarnations.

Although not the specific focus of this paper, a game-changing technology that is undoubtedly going to transform mobility is the advent of autonomous vehicles (AV). When AVs make their way into the market, it is plausible that some may choose to own autonomous vehicles in the traditional mode of private vehicle ownership, but others may eschew private vehicle ownership in favor of purchasing transportation by the trip or mile using shared autonomous vehicle fleets operated by transportation network companies that ply robo-taxis to meet on-demand mobility needs. Ride-sourcing and car-sharing usage is likely to grow further as autonomous vehicles are introduced in the marketplace.

Despite the interest in and growth of the new mobility services described above, there is a paucity of literature on the adoption, use, and impacts of these services. Travel

demand forecasting models, often used to support long-range transportation planning efforts, do not adequately account for the presence and growth of these mobility services, largely because of the lack of data about how consumers are using or intending to use (in the future) these services. There are complex interactions that affect how people may use different mobility-on-demand services. For example, high income individuals may find it convenient and affordable to use ride-sourcing services while they multi-task and use their travel time effectively (due to their high value of time). On the other hand, they may not use car-sharing services because they are more likely to own multiple cars (exhibit a higher level of private car ownership) and would rather multi-task and use travel time effectively rather than drive the car themselves. Understanding the complex interactions and competing or complementary forces that contribute to the use (or not) of mobility-on-demand services is critical to the development and enhancement of travel forecasting models for the future.

In this paper, a bivariate ordered probit model is specified and estimated on a survey data set derived from the 2014-2015 Puget Sound Regional Travel Study. This survey data set includes specific information about mobility-on-demand service usage (specifically, ride-sourcing and car-sharing services), in addition to the usual variables describing socio-economic, demographic, and activity-travel characteristics. The objective of the model development effort is to jointly model people's use of mobility-on-demand services, while accounting for common unobserved factors that may simultaneously affect people's proclivity to use these two services of interest.

The remainder of this paper is organized as follows. In the next section, a brief review of the literature is provided. The third section presents the modeling methodology, while the fourth section offers a description of the data set. Model estimation results are presented in the fifth section, elasticity computations are discussed in the sixth section, and concluding thoughts are offered in the seventh and final section.

CHAPTER 2 MOBILITY-AS-A-SERVICE AND TRAVEL DEMAND

The literature on the role, use, and impacts of mobility-as-a-service providers is rather sparse, largely because of the novelty of these services and the proprietary nature of data that makes it difficult to conduct extensive research on traveler behavior, values, and choices in relation to these platforms (Rayle et al., 2016). There is some aggregate data that illustrates the rapid adoption and growth of these services, but there is clearly a need for disaggregate choice modeling efforts that incorporate these services explicitly as choice options – thereby shedding light on the factors that affect their usage and the potential impacts they may have in the future.

Ride-sourcing services (such as Uber and Lyft) have experienced strong growth ever since they were introduced less than a decade ago. Uber is clearly the largest and most well-known service provider; the company began its operations in 2009. Lyft, which is a competitor to Uber, is relatively smaller; it began operations in 2012 and currently operates in about 220 cities, mostly concentrated in the United States and a few cities of Southeast Asia (Lyft, 2016). Both of these companies have been the beneficiaries of large investments by major companies; most notably, key auto manufacturers are partnering with these companies as they increasingly recognize the disruptive role that ride-sourcing services may play in the transportation eco-system, particularly in a future that will see the advent of autonomous vehicles (Buhr, 2016).

With a simple push of a smartphone app, it is possible for individuals to summon a ride, obtain door-to-door transportation, and pay for the service without having to engage in a physical monetary transaction. Because the services are regulated less than regular taxi companies, and ride-sourcing service drivers are simply driving their own vehicles to provide rides, it is possible for the services to expand rapidly and reach a large geographically dispersed market. As noted by Rayle et al. (2016), the overall impacts of ride-sourcing services on vehicle travel are unclear. When an individual uses ride-sourcing services to travel instead of using his or her own car, one vehicle trip is simply being replaced by another chauffeur-driven vehicle trip. The additional “empty-vehicle” miles involved in servicing the trip may add vehicle miles of travel (VMT). Rayle et al.

(2016) note that while ride-sourcing is being used in lieu of taxi in many instances, at least one-half of ride-sourcing usage replaced trips by modes other than taxi, including public transit and driving.

The rapid growth of ride-sourcing services has led to strong opposition from the taxi industry which feels that the mobility-on-demand services play by a different set of rules. The taxi industry is experiencing revenue and usage reductions worldwide (Waheed et al., 2015), making it difficult for taxi drivers to sustain their livelihood. In New York alone, for example, Uber served 93 million trips between April and September 2015; during the same period, regular taxi companies served about 88.4 million trips (Bialik et al., 2015). Although the use of ride-sourcing services is increasing, a review of the literature shows that very little is known about the socio-economic and demographic profile, and activity-travel characteristics, of ride-sourcing service users. Rayle et al. (2016) provide some data on ride-sourcing service users in comparison to regular taxi users in San Francisco. They find that ride-sourcing users are generally younger males who are highly educated and reside in zero-car households.

A larger body of literature can be found about car-sharing services, presumably because these services have been around for a longer period of time. Shaheen et al. (2009) identify three phases of the car-sharing industry, including, initial market entry and experimentation (1994-2002), growth and market diversification (mid-2002 to late-2007), and finally commercial mainstreaming (late-2007 to present). It may be argued that car-sharing services are experiencing further change at this time as they embrace the smartphone platform and use the power of apps to greatly ease the use of their services. Using smartphone based apps, it is now possible for individuals using car-sharing services to find cars in the immediate vicinity and rent cars for short periods of time on-demand.

A number of studies have attempted to analyze the impacts of car-sharing on mobility. Baptista et al. (2014) note that car-sharing contributes to more efficient mobility with a lower level of car ownership and use among members, lower demand for parking capacity, and lower vehicle acquisition and use costs as members shed personally owned

vehicles in favor of using shared vehicles. In a survey conducted in Germany, Firnkorn and Müller (2011) found that more than one-quarter of the respondents would be willing to forego a vehicle purchase if car2go was offered permanently. However, it may be important for member users to have easy access to other modes of transportation as well (in the event that a shared car is not available) before they would be willing to shed personally owned vehicles (Firnkorn, 2012).

Clewlow (2016) conducted a study of car-sharing members to better understand their characteristics. The study finds that lower levels of vehicle ownership among car-sharing members are limited to urban areas; however, suburban car-share members drive less than their non-car-share counterparts. The cars that car-share members do own are more likely to be alternative-fueled vehicles, thus suggesting that car-share members are more environmentally conscious, a finding also reported by Costain et al. (2012). Coll et al. (2014) found that socio-economic characteristics, namely, education, family structure, and non-motorized mode use, are strong predictors of car-sharing membership. Efthymiou et al. (2013) find that Greeks of low to medium income, who are environmentally conscious and take taxis for social reasons, have a higher probability of joining a car-share program. Car-sharing members derived the greatest utility from the service when a vehicle was available at their desired time and location (Zoepf and Keith, 2016), but showed some willingness to adjust trip timing and access distance in response to car availability.

Overall, it can be seen that there is widespread interest in understanding the role, use, and impacts of these disruptive transportation services, but there is very limited knowledge about these mobility platforms despite their growth in the transportation ecosystem. This study aims to fill this critical gap in the literature recognizing that both car-sharing and ride-sourcing services are increasingly leveraging technology, ubiquitous connectivity, and location-aware platforms to solve mobility challenges. Despite the potential relationship (whether synergistic, competing, or both, depending on the circumstances) between these two types of services, past studies have focused on one or the other without an integrated examination of both mobility services. Because both

services are technology enabled and involve relying on vehicles not owned privately by the individual, there are likely to be underlying unobserved factors that affect usage of both of these services. For example, people who embrace technology and enjoy using on-demand services facilitated by a smartphone app may be more likely to adopt disruptive transportation services (i.e., both ride-sourcing and car-sharing services). Technology savviness is then an unobserved lifestyle characteristic that affects service usage, thus necessitating a joint examination of the factors contributing to the use of mobility-on-demand services. This paper adopts a novel and more integrated approach to modeling the use of disruptive mobility services to address this need.

CHAPTER 3 DATA DESCRIPTION

Data for this study is derived from the 2015 household travel survey of the Puget Sound Regional Travel Study (PSRC, 2015). The survey was part of a larger regional travel study that aimed to obtain detailed information about the socio-economic, demographic, and activity-travel characteristics of a representative sample of the population in the region. In addition to obtaining information about socio-economic and activity-travel characteristics, the survey also collected data about attitudes and values, technology ownership and usage, membership in and use of car-sharing and ride-sourcing services, and future intended adoption and use of autonomous vehicle technologies. Thus, the survey data set includes a rich set of information conducive to analyzing the potential adoption and use of disruptive mobility platforms.

In order to prepare the sample for analysis, individuals under the age of 18 years were removed so that the modeling effort would be exclusively focused on the adult subsample. In addition, any record that involved proxy reporting was also removed because it was deemed potentially challenging for an individual to accurately report car-sharing and ride-sourcing usage on behalf of somebody else in the household. The final, filtered sample used for analysis and modeling comprised 2,789 adults. Respondents provided information on the frequency of ride-sourcing and car-sharing usage using the following categories (for each of the services):

- I never do this
- I do this, but not in the past 30 days
- I do this 1-3 times per month
- I do this 1 day per week
- I do this 2 or more days per week

As the categories represent an increasing level of usage, it was considered appropriate to treat the choice frequency of usage as an ordinal dependent variable.

Table 1 presents a summary of the sample characteristics. The age distribution shows that 21 percent of the sample is under 34 years of age and one-quarter of the sample is 65 years or over. Females constitute 57 percent of the sample, while those with a driver's license constitute 93 percent of the sample. It is found that a good majority of

the respondents (at 69 percent) own a smartphone. The income distribution shows that 34 percent of the individuals reside in households making \$100,000 or more per year. The sample is relatively well-educated, with 38 percent indicating that they have a Bachelor's degree and another 30 percent indicating that they have a graduate degree. A majority of the sample is employed, with 55 percent indicating either full-time (46 percent) or part-time (9 percent) employment. Just about one-third of the sample resides in single-person households; however, only 19 percent of the respondents indicate that they reside in households with children (a child is defined as an individual below the age of 18). Thus it appears that there is a large number of households with only multiple adults. It is found that 11 percent of the sample resides in households with no vehicles, 39 percent reside in households with one vehicle, and 35 percent reside in households with two vehicles.

With respect to the dependent variables themselves, it appears that the vast majority of respondents have never used mobility-on-demand services, thus suggesting that these services are still a novelty and are therefore likely to have only a minimal impact on travel demand and network performance at the current time. It is found that 86 percent never used ride-sourcing and 92 percent never experienced car-sharing service. Only about two percent of the sample indicates a regular usage (one day per week or more) of either disruptive mobility service. Between the two services, it can be seen that ride-sourcing usage is slightly higher than car-sharing usage. The characteristics of the sample, and the unique variables that provide a measure of mobility-on-demand service usage, make this data set ideally suited for this study.

Table 1 Survey Sample Description

Variable	Count	%	Variable	Count	%
Age			Education level		
18-24	87	3	Less than high school	36	1
25-34	511	18	High school graduate	171	6
35-44	461	17	Some college	424	15
45-54	452	16	Vocational/technical training	97	3
55-64	603	22	Associates degree	180	6
65-74	447	16	Bachelor degree	1053	38
75-84	185	7	Graduate/post-graduate degree	828	30
85 or older	43	2	Employment Status		
Gender			Employed full-time (paid)	1295	46
Male	1197	43	Employed part-time (paid)	240	9
Female	1592	57	Self-employed	173	6
Has a smartphone			Unpaid volunteer or intern	34	1
Yes	1927	69	Homemaker	145	5
No	862	31	Retired	718	26
Has a valid driver's license			Not currently employed	184	7
Yes	2603	93	Single person household		
No	186	7	No	1860	67
Residential density			Yes	929	33
Up to 5,000 hh per square mile	1994	71	Has children in household		
Above 5,000 hh per square mile	795	29	No	2268	81
Annual household income			Yes	521	19
Under \$25,000	355	13	Household vehicle count		
\$25,000-\$49,999	575	21	0 (no vehicles)	316	11
\$50,000-\$74,999	483	17	1	1099	39
\$75,000-\$99,999	430	15	2	986	35
\$100,000 or more	946	34	3	268	10
Ride-sourcing Frequency (last 30 days)			4	77	3
I never do this	2386	86	5 or more	43	2
I do this, but not in the past 30 days	170	6	Car-sharing Frequency (last 30 days)		
I did this 1-3 times in the past 30 days	171	6	I never do this	2529	92
I did this 1 day per week	33	1	I do this, but not in the past 30 days	122	4
I did this 2 or more days per week	29	1	I did this 1-3 times in the past 30 days	97	4
			I did this 1 day per week	18	1
			I did this 2 or more days per week	23	1

CHAPTER 4 MODELLING METHODOLOGY

This study involves the joint modeling of two ordinal dependent variables, with the possible presence of common unobserved factors (such as attitudes or lifestyle preferences) that affect both the usage of ride-sourcing and car-sharing services. For this reason, a bivariate ordered probit modeling methodology is adopted in this study. The multivariate probit modeling methodology has been used in several studies in the travel behavior modeling domain and elsewhere (e.g., Emmerink et al., 1996 and Ferdous et al., 2010). The bivariate ordered probit stitches together two ordered probit equations while accommodating error covariance that may exist between them. The correlated error terms are assumed to follow a bivariate normal distribution and the model parameters may be estimated using maximum likelihood estimation methods. The model is estimated (and may be applied) at the person level, and not at the individual trip level. The model is therefore not akin to a traditional mode choice model that may include explanatory variables such as trip time or cost; rather, it is a person level model that purports to shed light on the potential adoption and intensity (frequency) of use of each of the disruptive transportation services while accounting for unobserved lifestyle preferences that may affect their use. This section offers a brief overview of the modeling methodology and formulation.

Assume that there are underlying continuous latent variables whose partitioning directly relates to the frequency of use of ride-sourcing and car-sharing services. Let q be an index for observation units (in this case individuals) ($q = 1, 2, \dots, Q$). Let f_q and g_q represent the frequency categories of usage of ride-sourcing and car-sharing services, respectively. Also, let m and n be, respectively, indices for the discrete outcomes corresponding to the frequency categories of ride-sourcing and car-sharing. This means that m and n may take the values of “I never do this” ($m = 1$ or $n = 1$), “I do this, but not in the past 30 days” ($m = 2$ or $n = 2$), and so on until “I did this 2 or more days per week” ($m = 5 = M$ or $n = 5 = N$). The model takes the following form:

$$\begin{aligned} f_q^* &= \alpha' x_q + v_q & \text{where } f_q &= m \text{ if } \delta_{m-1} < f_q^* < \delta_m, \delta_0 = -\infty, \delta_M = \infty \\ g_q^* &= \beta' y_q + \eta_q & \text{where } g_q &= n \text{ if } \psi_{n-1} < g_q^* < \psi_n, \psi_0 = -\infty, \psi_N = \infty \end{aligned} \quad (1)$$

In the above equation, f_q^* and g_q^* are the latent variables for individual q . They indicate the propensity of an individual to use ride-sourcing and car-sharing services, respectively. The larger the latent variable, the greater the frequency of usage; x_q and y_q are vectors containing all exogenous covariates of the model for individual q that affect the latent variables (with no constant term); α and β are vectors of the coefficients to be estimated, and which capture the effects of the exogenous variables x_q and y_q ; δ_m and ψ_n are the thresholds that partition the latent variable into the same number of segments as there are categories; m and n indicate the categories of the dependent variables, and M and N indicate the total number of categories for each of the dependent variables; and v_q and η_q are the random error terms of the latent variable equations.

In the current study, normal marginal distributions are assumed for these error terms. The error terms v_q are assumed to be independent and identically distributed (IID) across individuals q , and the error terms η_q are also assumed to be IID across individuals q . Furthermore, a standard normal distribution is used for the error terms (these standardizations are innocuous normalizations needed for econometric identification). In addition, to accommodate for the potential presence of correlation in these error terms (due to unobserved factors such as technology savviness, availability of disposable income, and openness to shared-economy services), a joint bivariate standard normal distribution is considered for the error terms v_q and η_q .

The parameters to be estimated in the joint bivariate ordered response model include the α and β vectors, the $M-1$ δ_m parameters ($\delta_0 = -\infty, \delta_M = \infty, -\infty < \delta_1 < \delta_2 < \dots < \delta_{M-1} < \infty$), the $N-1$ ψ_n parameters ($\psi_0 = -\infty, \psi_N = \infty, -\infty < \psi_1 < \psi_2 < \dots < \psi_{N-1} < \infty$), and the θ parameter characterizing the correlation between the error terms. To write the log-likelihood function, define $I_q(m, n)$ as a binary indicator variable that takes the value of 1 if individual q falls in frequency category m for ride-sourcing service use and frequency category n for car-sharing service use, and 0 otherwise, and

$\Pr[f_q = m, g_q = n]$ as the probability of the occurrence $I_q(m, n) = 1$. Then, the log likelihood function for the model takes the following form:

$$\log L = \sum_{q=1}^Q \sum_{m=1}^M \sum_{n=1}^N I_q(m, n) \log \Pr[f_q = m, g_q = n] \quad (2)$$

Also, let $b_{qm} = \delta_m - \alpha' x_q$ and $d_{qn} = \psi_n - \beta' y_q$. Then, the probability of the occurrence is:

$$\begin{aligned} \Pr[f_q = m, g_q = n] &= \Pr[\delta_{m-1} < f_q^* < \delta_m \text{ and } \psi_{n-1} < g_q^* < \psi_n] \\ &= \Pr[b_{q,m-1} < v_q < b_{qm} \text{ and } d_{q,n-1} < \eta_q < d_{qn}] \\ &= \Phi_2[b_{qm}, d_{qn}; \theta] - \Phi_2[b_{q,m-1}, d_{qn}; \theta] - \Phi_2[b_{qm}, d_{q,n-1}; \theta] \\ &\quad + \Phi_2[b_{q,m-1}, d_{q,n-1}; \theta] \end{aligned} \quad (3)$$

where Φ_2 is the bivariate cumulative normal distribution function. All of the parameters in the model are estimated by maximizing the log-likelihood function above using the GAUSS matrix programming language.

CHAPTER 5 MODEL ESTIMATION RESULTS

This section presents a detailed discussion of the model estimation results, which are shown in Table 2. The final model, chosen after testing numerous alternative specifications, includes a number of socio-economic characteristics as well as interaction effects that provide deeper insights on differences among individuals in the propensity to use ride-sourcing and car-sharing services.

At the outset, it should be noted that despite a specification that includes a number of exogenous variables, the error correlation of 0.401 is highly statistically significant (t-statistic of 8.725). The presence of this significant correlation justifies the use of the bivariate ordered probit model formulation and suggests that travel forecasting models need to recognize the potential presence of such factors to more accurately assess and predict the impacts of disruptive transportation services on travel demand and network performance. It should also be noted that the correlation is positive indicating that unobserved factors that positively contribute to the use of one disruptive service (say, ride-sourcing) also contribute positively to the use of the other disruptive service (say, car-sharing). This is consistent with expectations; for example, technology-savvy individuals who embrace a technology-driven lifestyle or individuals who are more adventurous and risk-taking in nature are likely to adopt and use both services.

Individuals with a higher education level exhibit a greater propensity to use both ride-sourcing and car-sharing services. It is likely that these individuals are more likely to be aware of such services, and have the ability to leverage such services through the use of technology. Similar findings have been reported by Rayle et al. (2016), Viechnicki et al. (2015), Martin and Shaheen (2011), and Coll et al. (2014). Older individuals, on the other hand, are less likely to use ride-sourcing and car-sharing services, presumably because they are not as adept at using technology and trying new services as younger individuals. However, it is plausible that an aging population will increasingly adopt and use ride-sourcing services as their ability to drive diminishes over time.

Those who are employed full-time or self-employed exhibit a greater propensity to use ride-sourcing and car-sharing services relative to those who are unemployed or

employed part-time. This effect is found even after controlling for income, suggesting that employed individuals may be using the services because of work-related activities. Not having a driver's license is a natural deterrent to car-sharing usage, and this is evidenced by the statistically significant negative coefficient. Ride-sourcing usage, on the other hand, does not show a difference based on driver's license holding status.

The next set of variables correspond to the effects of smartphone ownership and single person household. The pattern of the coefficients indicates, as expected, that those who have a smartphone exhibit a greater propensity to use both ride-sourcing and car-sharing services. The magnitude of the smartphone effect is larger for ride-sourcing services, which is consistent with expectations because a smartphone app is the only way to use such services, but is often not absolutely required to avail car-sharing services. Further, while there is no effect of family structure on ride-sourcing propensity, the results indicate that individuals who are single and do not own a smart phone are the least likely to car-share. Also, compared to non-single individuals, single individuals, when they own smartphones, have a higher propensity to car-share than when they do not own smartphones. This may be attributable to single individuals being generally more socially active, and using smartphones as a means to securing a more private travel arrangement through car-sharing when they want to drive together with friends and significant others.

There is an observed income effect, showing that lower income individuals have lower propensities to use both ride-sourcing and car-sharing services, likely due to cost considerations. This effect of low income is further amplified when children are present, possibly because of tightening budgets when children are present. This interpretation is supported by the fact that, when children are present, even the middle income category (\$50K-\$100K) sees a reduction in both ride-sourcing and car-sharing propensities. In addition, households with children may undertake more complex tours and more activities in general, thus rendering it more challenging to accomplish all activities and trips in a cost-economical manner using ride-sourcing and car-sharing services. Individuals in the highest income category (regardless of whether they have or do not have children) and individuals in the middle income category with no children have the

highest ride-sourcing and car-sharing propensities relative to individuals from other income-family structure groups. In general, recent surveys have shown that sharing-economy based services are largely used by the young, rich, and well-educated (e.g., Smith, 2016), and the model estimation effort of this paper offers very consistent findings.

The results in Table 2 also show that individuals in households with more vehicles have lower propensities to use ride-sourcing services, but only if living in low density (≤ 5000 households per square mile) neighborhoods (the coefficient on one vehicle is -0.673, and the coefficient on 2+ vehicles is -0.908, in low density areas, but the effective coefficients become zero for both one vehicle households and 2+ vehicle households in high density (> 5000 households per square mile) areas because of the interaction terms). Also, there is no difference in ride-sourcing propensity based on residential living for zero vehicle households. As to car-sharing propensity, individuals in households with more vehicles are less likely to car-share; this is consistent with expectations as individuals would probably find it more convenient and cost-effective to just drive one of their own household vehicles rather than a car-sharing vehicle. Prior literature (Clewlow, 2016; Coll et al., 2014) has reported similar results. When taken together with the interaction terms of number of vehicles with residential density of living (for the car-sharing service), the implication is that the negative effect of vehicle ownership on car-sharing is less pronounced in dense areas (note that the coefficient on one vehicle is $-1.292 + 0.300 = -0.992$ in dense areas relative to -1.292 in non-dense areas, and on two vehicles is $-2.042 + 0.754 = -1.288$ in dense areas relative to -2.042 in non-dense areas). Another perspective on this is that individuals from one vehicle and two vehicle households living in high density areas have a higher car-sharing propensity than their peers from one vehicle and two vehicle households living in relatively low density areas. However, as one would expect, there is no difference between individuals in low density and high density neighborhoods in their car-sharing propensity if they do not own any vehicles. Overall, households who own vehicles in *dense* residential areas have higher propensities for ride-sourcing and car-sharing services relative to their peers in

less dense residential areas. Thus density of development and access to destinations appear to play a role in shaping usage of these services. Dense areas may also have higher parking costs that motivate individuals to use ride-sourcing and car-sharing services.

The model shows an acceptable fit, as indicated by the difference between the log-likelihood of the null model (i.e., the model with only the thresholds in each of the ordered probit equations and an assumption that the error correlation is zero) and the log-likelihood of the full model. Overall, it can be seen that car-sharing and ride-sourcing service users tend to be young, urban residents who are rich and well-educated. Although these findings have been documented in prior studies (Smith, 2016), this is the first study that actually quantifies the effects of different variables (and their interactions) on the propensity to use ride-sourcing and car-sharing services. The study offers a model that may be integrated within overall travel forecasting model systems with a view to better predict the adoption, use, and impacts of such services in the future.

Table 2 Estimation Results for Bivariate Ordered Probit Model

(coefficients represent impact of variables on underlying propensities of ride-sourcing and car-sharing)

Variable	Ride-sourcing		Car-sharing	
	Coef	t-stat	Coef	t-stat
Education (Base: Not achieved Bachelor's degree level)				
Bachelor's degree or higher	0.326	3.554	0.380	3.474
Age (Base: 18-34 years)				
35-54 years	-0.419	-5.617	0.000	-
55 years and above	-1.113	-10.976	-0.408	-4.293
Employment (Base: Unemployed)				
Employed Part time	0.000	-	0.000	-
Employed (Full time)	0.199	2.348	0.242	2.5
Employed (Self-employed)	0.199	2.348	0.242	2.5
Valid driver's license ownership (Base: No)				
Yes	0.000	-	1.551	6.066
Smartphone Ownership & Family Structure (Base: Doesn't have a Smartphone & Multi person HH)				
No smartphone and a single person HH	0.000	-	-0.387	-1.577
Has a Smartphone x Non-single person HH	1.133	7.275	0.476	3.013
Has a Smartphone x Single person HH	1.133	7.275	0.476	3.013
Income (Base: Above \$100,000)				
Below \$49,999	-0.272	-3.144	-0.185	-1.735
Below \$49,999 x Presence of children	-1.281	-3.566	-0.758	-3.260
\$50,000-\$99,999 x Presence of children	-0.680	-4.113	-0.943	-4.565
Vehicle Ownership (Base: 0 Vehicles)				
1 vehicle	-0.673	-7.145	-1.292	-10.326
2 or more vehicles	-0.908	-11.357	-2.042	-14.609
1 vehicle x high density living	0.673	7.145	0.300	2.68
2+ vehicles x high density living	0.908	11.357	0.754	5.604
Threshold Values				
δ_1 and ψ_1	0.172	1.379	-0.025	-0.142
δ_2 and ψ_2	0.623	4.978	0.415	2.350
δ_3 and ψ_3	1.455	11.117	1.097	5.993
δ_4 and ψ_4	1.826	13.827	1.379	7.203

Error correlation = 0.401 (t-stat: 8.725)

Log-likelihood (Null model): -3136.258

Log-likelihood (Full model): -2117.166

Pseudo-R² (McFadden) = 0.325

Notes: 1. The "Unemployed" category groups the following categories from Table 1: "Unpaid volunteer or intern", "Homemaker", "Retired" and "Not currently employed".

2. Null model is the model with only the thresholds in each of the ordered probits (sample shares model) with the constraint that the correlation term is zero.

CHAPTER 6 COMPUTATION OF PSEUDO-ELASTICITIES

The parameters on the exogenous variables in Table 2 do not directly provide a sense of the absolute magnitude of the effects of variables. One can obtain the elasticity effects of each variable on each of the 25 bivariate combination levels of ride-sourcing and car-sharing to capture the correlations between the levels of ride-sourcing and car-sharing. But, for presentation ease, in this paper, only the separate effects of variables on the univariate marginal levels of ride-sourcing and car-sharing are considered. Further, to facilitate an understanding of the order-of-magnitude effects of variables, cardinal values are assigned to each of the ordinal levels of ride-sourcing and car-sharing. Next, “pseudo-elasticity” effects of exogenous variables on the expected total number of instances per month of each of ride-sourcing and car-sharing use are computed. For these computations, it is assumed that an individual uses these services no more than once a day. The cardinal value assignments for the ordinal frequency levels in the model are as follows: (1) I never do this = 0 instances per month, (2) I do this, but not in the past 30 days = 0.333 instances per month, (3) I did this 1-3 times in the past 30 days = 2 instances per month, (4) I did this one day per week = 4 instances per month, (5) I did this two or more days per week (say 4 instances per week) = 16 instances per month. With these assignments, and using the notation c_m for the cardinal value assignment corresponding to ride-sourcing level m , the marginal expected value of the frequency of ride-sourcing use per month for individual q (\tilde{f}_q) is:

$$E(\tilde{f}_q) = \sum_{m=1}^5 c_m \times \Pr[f_q = m] \quad (4)$$

Similarly, using the notation d_n for the cardinal value assignment corresponding to car-sharing level n , the marginal expected value of the frequency of car-sharing use per month for individual q (\tilde{g}_q) is:

$$E(\tilde{g}_q) = \sum_{n=1}^5 d_n \times \Pr[g_q = n] \quad (5)$$

The equations above may be used to compute the aggregate-level “pseudo-elasticity effects” of exogenous variables. For variables that have an interaction effect with another variable, the elasticities are computed for all sub-groups characterized by the main and interaction effects. For example, for car-sharing, consider the main effects of smartphone ownership and the interaction effect with whether the individual belongs to a single-person household. To examine the joint effects of smartphone ownership and family structure, four multinomial sub-groups are developed: (1) no smartphone ownership, non-single, (2) no smartphone ownership, single, (3) smartphone ownership, non-single, and (4) smartphone ownership, single. The elasticity effects are computed for these variables with respect to the first sub-group (no smartphone ownership, non-single) as the base instance. Based on the results in Table 2, there is no ride-sourcing propensity difference between the first two sub-groups listed above; so, in this case, both of the first two sub-groups constitute the base. Also, for the ride-sourcing frequency, there are no interaction effects; so the elasticity effects for the last two sub-groups should be the same as well. A similar approach is adopted for other cases of interaction effects.

Additional details on the computation of aggregate level elasticities are available in Dias et al. (2016) and are omitted here in the interest of brevity. Table 3 provides the computed pseudo-elasticity effects. The first entry in the table (corresponding to education) indicates that the number of instances of ride-sourcing use per month for individuals who hold a Bachelor’s degree or higher education level is, on average, about 32% more than the number of instances of ride-sourcing use per month for individuals who do not have a Bachelor’s degree. Other entries may be similarly interpreted. These elasticity effects indicate the particularly strong negative effects of older age (55 years or above) and the presence of children in non-high income households on both ride-sourcing and car-sharing propensity. Low density residential living has a substantial negative impact (relative to high density residential living) on ride-sharing propensity in non-zero

vehicle households, or, conversely, high density residential living has a substantial positive impact (relative to low density residential living) on ride-sharing propensity in non-zero vehicle households, though there is not much of a residential density effect for car-sharing propensity. Also, while the count of private motorized vehicles has a negative impact on ride-sourcing in low density areas, there is no such effect of vehicle ownership in high density areas. On the other hand, high vehicle ownership levels have a definite negative impact on car-sharing regardless of residential density. Not surprisingly, among the strongest facilitators of ride-sourcing and car-sharing is the ownership of smartphones.

Table 3 Computation of Pseudo-Elasticity Effects

Variable	Ride-sourcing	Car-sharing	Both
Education (<i>Base: Not achieved Bachelor's degree level</i>)			
Bachelor's degree or higher	32.3%	38.9%	35.4%
Age (<i>Base: 18-34 years</i>)			
35-54 years	-38.7%	0.0%	-24.2%
55 years and above	-75.7%	-41.2%	-62.7%
Employment (<i>Base: Unemployed</i>)			
Employed (Part-time)	0.0%	0.0%	0.0%
Employed (Full time)	26.5%	36.4%	30.9%
Employed (Self-employed)	26.5%	36.4%	30.9%
Valid driver's license ownership (<i>Base: No</i>)			
Yes	0.0%	868.5%	76.4%
Smartphone Ownership & Family Structure (<i>Base: Doesn't have a Smartphone & Multi person HH</i>)			
Doesn't have a Smartphone & Single person HH	0.0%	-42.7%	-28.8%
Has a Smartphone & Multi person HH	355.9%	87.0%	174.8%
Has a Smartphone & Single person HH	355.9%	87.0%	174.8%
Income & Children in Household (<i>Base: Above \$100,000 & No Kids</i>)			
Below \$49,999 & No kids	-27.5%	-20.9%	-24.5%
Below \$49,999 & Has kids	-88.0%	-72.7%	-81.0%
\$50,000-\$99,999 & No kids	0.0%	0.0%	0.0%
\$50,000-\$99,999 & Has kids in HH	-55.7%	-72.7%	-63.4%
Above \$100,000 & Has kids in HH	0.0%	0.0%	0.0%
Vehicle Ownership & Residential Density (<i>Base: 0 Vehicles & Non-dense region</i>)			
0 vehicles & Dense	0.0%	0.0%	0.0%
1 Vehicle & Non-dense	-54.9%	-76.7%	-70.0%
1 Vehicle & Dense	0.0%	-65.7%	-45.5%
2 Vehicles & Non-dense	-66.9%	-92.3%	-84.5%
2 Vehicles & Dense	0.0%	-76.6%	-53.0%

CHAPTER 7 CONCLUSIONS

Ride-sourcing and car-sharing are two disruptive transportation services whose adoption, use, and impacts in the marketplace remain poorly understood despite their proliferation. Both car-sharing and ride-sourcing services are increasingly prevalent in many cities around the world. Forecasting the impacts of these transformative transportation technologies on activity-travel demand, network performance, and land use development patterns requires a clear understanding of the factors that contribute to the use of these services and the development of models capable of predicting market adoption patterns for different socio-economic groups in a wide variety of contexts. The development of such an understanding and appropriate predictive model systems has been stymied by the lack of disaggregate behavioral data on the adoption and use of these services.

This paper aims to address this gap in the literature by jointly modeling the propensity of individuals to use ride-sourcing and car-sharing services as a function of numerous exogenous variables including socio-economic and demographic variables, smartphone ownership, and density of residential location. The model is estimated on a survey sample of 2,789 adults who participated in the 2014-2015 Puget Sound Regional Travel Study. The survey data set included two questions regarding the frequency of use (over a 30-day period) of ride-sourcing services and car-sharing services. These two variables were treated as ordinal dependent variables, and a number of exogenous variables were used to explain disruptive mobility service usage. In general, it was found that users of these services tend to be young, well-educated, higher-income, employed, and residing in higher density neighborhoods. These findings are consistent with those found in other studies. However, while past studies have largely presented descriptive statistics, this paper presents a model that can be used to infer and quantify the effects of various exogenous factors on the usage of mobility-on-demand services. It was found that there are significant interaction effects that explain ride-sourcing and car-sharing usage; for example, the presence of children appears to reduce ride-sourcing and car-sharing usage among low and middle-income households, possibly due to budget constraints and more complex activity-travel patterns that the presence of children engenders. Similarly,

it was found that households with vehicles are less likely to use car-sharing services; however, households with one or more vehicles and residing in a high-density location are more likely than their peers residing in low density areas to use both ride-sourcing and car-sharing services – suggesting that not all vehicle-owning households are created equal. Also, given that most households in the US own one or more motorized vehicles, the results suggest that neighborhood densification is a strategy that would contribute to greater usage of ride-sourcing services. At the same time, the study results underscore the importance of treating these new mobility services as separate transportation modes in land use-transportation planning. For example, densification may lead to less private vehicle use, but increased ride-sourcing demand, negating some of the potential VMT reduction benefits associated with neo-urbanist designs. Ride-sourcing may also draw away from walk, bicycling, and public transportation modes in high density areas, further contributing to increases in VMT and greenhouse gas emissions.

The model presented in this paper may be used to inform travel forecasts that are currently limited with respect to their ability to reflect usage of disruptive mobility-on-demand services. The model reflects the effects of different variables on frequency of service usage, thus allowing the prediction of market adoption and use among various socio-economic groups under alternative future scenarios. The model presented in this study is at the person-level, and hence the model could not account for supply side attributes (cost and travel time) in predicting service usage. Nevertheless, the model may be used to predict longer term strategic travel choices that are made at the agent-level (similar to vehicle ownership); such choices play an important role in day-to-day tour or trip level decisions. Future modeling efforts could focus on predicting mobility service usage at the tour or trip level while incorporating supply side attributes in addition to unobserved lifestyle factors. Travel surveys should explicitly represent mobility services as distinct modes so that such tour- and trip-level models can be estimated. The development of random parameter models that can better reflect heterogeneity in the population may also be a worthy endeavor.

APPENDIX A: GAUSS CODE FOR MODEL ESTIMATION

```
library maxlik,cml;
#include maxlik.ext;
maxset;
#include cml.ext;
cmlset;

/* Correlated OR system */

clearg nbin,ncont1,ncont2,cop1,cop2,nc1,nc2,nthresh1,nthresh2;

/* Number of rows for likelihood function to read */

__row = 2789;

/* Data set to be used */

dataset = "C:\\Felipe\\Puget_MNP\\Gauss_Files\\Puget_Data.dat";

/* Copula type; 1-normal, 2-FGM, 3-Frank, 4-Gumbel, 5-Clayton, 6-Joe */
/* Copula type for OR regimes */
cop1 = 1;

//Number of ordered categories in the first OR system
nc1 = 5;

//Number of ordered categories in the second OR system
nc2 = 5;

inf= 200;
minf=-200;
```

The GAUSS code presented here was originally developed by Dr. Chandra R. Bhat.

```

/*****
Specification of variables area
*****/
// First OR: frequency of ridesourcing
let ivcont1 = {
EMPNT_13
SPHN_2_3
RESDENS_2
AGE_7_8
AGE_9P
INC12
VEH1
VEH2P
edsing
edveh2
emsing
rdveh1
rdveh2
rdag78
rdag9p
kdin12
kdin34
};

// Second OR: frequency of carsharing
let ivcont2 = {
EDUC_6_7
EMPNT_13
GENDER_1
LICEN_2
SPHN_2_3
AGE_9P
INC12
VEH1
VEH2P
gnag9p
spsing
emsing
rdveh1
rdveh2
kdinc14
};
//first OR dependent variable
{ vvbin,dvcont1 } = indices(dataset,"ridefreq");

//second OR dependent variable
{ vvcont,dvcont2 } = indices(dataset,"carsfreq");

{ varcont1,ivtcont1 } = indices(dataset,ivcont1');
{ varcont2,ivtcont2 } = indices(dataset,ivcont2');

```



```

/* Create a matrix of ones and zeros to define constraints across the two OR
regimes,
the last columns is for correlation parameter
EQMAT should have as many columns as rows(ivtcont1)+rows(ivtcont2)+1, and as
many rows as the
number of distinct parameters */

EQMAT = eye(rows(ivtcont1)+rows(ivtcont2));

if cop1 == 1 or cop1 == 2 or cop1==7;
    brho1 = 0.1;
elseif cop1 == 3 ;
    brho1 = 1;
elseif cop1==4 or cop1 == 6;
    brho1 = 1.1;
elseif cop1==5;
    brho1 = 0.01;
endif;

//threshold counters
nthresh1 = nc1 - 1;
nthresh2 = nc2 - 1;

//thresholds
bthresh1 = -0.2948 | 0.1628 | 1.0147 | 1.3952 ;
bthresh2 = 0.1607 | 0.6087 | 1.2945 | 1.5730 ;

//insert only the coeffients corresponding to the ordered propensity equations

bstart =
{
    0.2913
    -1.1697
    -0.3340
    -0.2488
    -0.8550
    -0.3650
    -0.7467
    -1.3254
    0.3341
    0.4352
    -0.3349
    0.5960
    0.7572
    -0.3259
    -0.5792
    -1.2721
    -0.6759
    0.3932
    0.3758

```

```

0.1918
-1.5569
-0.4217
-0.2158
-0.1890
-1.3268
-2.0562
-0.4946
-0.5620
-0.4100
0.3163
0.7487
-0.6116
};

bstart = bstart';
brho1 = 0.4127;

b = bthresh1|bthresh2|bstart|brho1;

/*****
Main Program area
*****/

/* associating columns with variable names for output */

pname4 = "rho";

pthresh1 = "A_1"|"A_2"|"A_3"|"A_4";
pthresh2 = "B_1"|"B_2"|"B_3"|"B_4";

pthresh = pthresh1|pthresh2;

pname1 = 0 $+ "A_" $+ varcont1;
pname2 = 0 $+ "B_" $+ varcont2;

ncont1 = rows(pname1);
ncont2 = rows(pname2);

/* Change command below if you are using constraints */
_max_ParNames=pthresh|pname1|pname2|pname4;

/* _max_GradProc = &lgs; */
_max_Options = { bfgs stepbt };
_max_CovPar = 2;
_max_gradtol = 5e-04;
//_max_maxIters = 0;

// When running the copula regression, replace the final value with 1
// When running the copula without the correlation, replace the final value
with 0

```

```

//_max_active = ones(rows(bthresh1)+rows(bthresh2),1)|ones(rows(eqmat),1)|0;
_max_active = ones(rows(bthresh1)+rows(bthresh2),1)|ones(rows(eqmat),1)|1;

//_max_active      =      0|ones(rows(bthresh1)-1,1)|0|ones(rows(bthresh2)-
1,1)|ones(rows(eqmat),1)|1;

//print rows(_max_active)~rows(b);

__title = "COPULA-BASED correlated Ordered Response System ESTIMATION";
print "Copula type; 1 - normal, 2- FGM, 3 - Frank, 4 - Gumbel, 5 - Clayton, 6-
Joe";
print cop1;
print;

{ x,f,g,cov,retcode } = maxprt(maxlik(dataset,0,&lpr,b));


    //_cml_ParNames = pname1|pname2|pname3|pname4|"sig1"|"sig2";
_cml_ParNames = pthresh|pname1|pname2|pname4;
//print rows(_cml_ParNames);
_cml_Options = { bfgs stepbt };
_cml_CovPar = 2;
_cml_active = ones(rows(eqmat)-1,1)|1;


boun = { -100 100 };
boun = ones(nthresh1 + ncont1 + nthresh2 + ncont2,1).*boun;
//print rows(boun)~rows(boun');

if cop1 == 1 or cop1 == 2 or  cop1==7;
    boun1 = { -0.99 0.99};
elseif cop1 == 3 ;
    boun1 = { -100 100};
elseif cop1==4 or cop1 == 6;
    boun1 = { 1.00 100 };
elseif cop1==5;
    boun1 = { 0.01 100 };
endif;

bounsig = { 0.01 100,
            0.01 100 };

_cml_bounds = boun|boun1;

//print _cml_bounds;
//{ x,f,g,cov,retcode } = CMLprt(CML(dataset,0,&lpr,b));

print;

```

```

print;
print "Parameter Estimate St. Error T-stat ";
print;

out1 = x;
out2=(sqrt(diag(cov)));
out3=(x./(sqrt(diag(cov))));

omat = _cml_parnames~x~(sqrt(diag(cov)))~(x./(sqrt(diag(cov))));

mask = 0~1~1~1; // mask to print first column as strings
let fmt[4,3]=
    "-.*.*s " 8 8 @ format info for variable names column @
    ".*.*1f" 10 4 @ format info for mean column @
    ".*.*1f" 10 4 @ format info for mean column @
    ".*.*1f" 10 4 @ format info for mean column @;

printfm(omat,mask,fmt);

output off;

/*****
Procedures
*****/

/* Procedure for log likelihood function calculation */

proc lpr(x,dta);
    local ydb,ydc,yib,xib,yic1,xic1,yic2,xic2,xrho,xsig,y,c1,c2,xrhoo,
    c11,c12,F1,F2,F3,df2,df3,d,r11,r21,r12,s11,s21,s12,r,s,rnew,snew,p1,p2,z,gg,
    xthresh1,xthresh2,atu1, atl1,btu1, bt11,aF1,aF2,bF1,bF2;

    x      =      x[1:nthresh1:nthresh2]|(EQMAT'*x[nthresh1:nthresh2+1:(rows(x)-
1)])|x[rows(x)];
    xthresh1 = x[1:nthresh1];
    xthresh2 = x[nthresh1+1:nthresh1:nthresh2];

    ydb = dta[.,dvcont1] + 1;
    ydc = dta[.,dvcont2] + 1;

    yic1 = dta[.,ivtcont1];
    xic1 = x[(nthresh1 + nthresh2 + 1):(nthresh1 + nthresh2 + ncont1)];

```

```

yic2 = dta[:,ivtcont2];
xic2 = x[(nthresh1 + nthresh2 + ncont1 + 1):(nthresh1 + nthresh2 + ncont1 +
ncont2)];

xrho = x[rows(x)];

c1 = yic1*xic1;
c2 = yic2*xic2;

//atu1 = (submat(minf|x[1:nthresh1]|inf,ydb+1,0))- c1;
//atl1 = (submat(minf|x[1:nthresh1]|inf,ydb,0)) - c1;

//btu1 = (submat(minf|xthresh2|inf,ydc+1,0))- c2;
//btl1 = (submat(minf|xthresh2|inf,ydc,0)) - c2;

atu1 = (submat(minf|x[1:nthresh1]|inf,ydb,0))- c1;
atl1 = (submat(minf|x[1:nthresh1]|inf,ydb-1,0)) - c1;

btu1 = (submat(minf|xthresh2|inf,ydc,0))- c2;
btl1 = (submat(minf|xthresh2|inf,ydc-1,0)) - c2;

if cop1 == 1 or cop1 == 2 or cop1==7;
    xrho[1]= -0.999.*(xrho[1].<=-1)+xrho[1].*(xrho[1].>-1 and xrho[1] .< 1)
+0.999.*(xrho[1].>=1);
elseif cop1==4 or cop1 == 6;
    xrho[1] = xrho[1].*(xrho[1].>1)+1.001.*(xrho[1] <= 1);
elseif cop1==5;
    xrho[1] = xrho[1].*(xrho[1].>0)+0.001.*(xrho[1] <= 0);
endif;

aF1 = cdfn(atu1);
aF2 = cdfn(atl1);

bF1 = cdfn(btu1);
bF2 = cdfn(btl1);

//print          ydb[1:10,.]~(aF1[1:10,.]-aF2[1:10,.]~ydc[1:10,.]~(bF1[1:10,.] -
bF2[1:10,.]));

if cop1 == 1;
    r = cdfbvn(cdfni(aF1),cdfni(bF1),xrho) - cdfbvn(cdfni(aF2),cdfni(bF1),xrho)
- cdfbvn(cdfni(aF1),cdfni(bF2),xrho) +
cdfbvn(cdfni(aF2),cdfni(bF2),xrho) ;

elseif cop1 == 2;
    r = aF1 .* bF1 .* (1 + xrho .* (1 - aF1) .* (1 - bF1)) - aF2 .* bF1 .*
(1 + xrho .* (1 - aF2) .* (1 - bF1))

```

```

        -aF1 .* bF2 .* (1 + xrho .* (1 - aF1) .* (1 - bF2)) + aF2 .* bF2 .*
(1 + xrho .* (1 - aF2) .* (1 - bF2));

elseif cop1 == 3;
    r = ln (1+ ((exp(-xrho .* aF1) - 1).*(exp(-xrho .* bF1) - 1)./ (exp(-
xrho)-1)))./ (-xrho) - ln (1+ ((exp(-xrho .* aF2) - 1).*(exp(-xrho .* bF1) -
1)./ (exp(-xrho)-1)))./ (-xrho)
        -ln (1+ ((exp(-xrho .* aF1) - 1).*(exp(-xrho .* bF2) - 1)./ (exp(-
xrho)-1)))./ (-xrho) + ln (1+ ((exp(-xrho .* aF2) - 1).*(exp(-xrho .* bF2) -
1)./ (exp(-xrho)-1)))./ (-xrho);

elseif cop1 == 4;
    r = exp(-((((-ln(aF1))^xrho + (-ln(bF1))^xrho)^(1 ./xrho)))) - exp(-((((-
ln(aF2))^xrho + (-ln(bF1))^xrho)^(1 ./xrho))))
        -exp(-((((-ln(aF1))^xrho + (-ln(bF2))^xrho)^(1 ./xrho)))) + exp(-((((-
ln(aF2))^xrho + (-ln(bF2))^xrho)^(1 ./xrho)))) ;

elseif cop1 == 5;
    r = ((aF1^(-xrho) + bF1^(-xrho) - 1)^(-1./xrho)) - ((aF2^(-xrho) + bF1^(-
xrho) - 1)^(-1./xrho))
        -((aF1^(-xrho) + bF2^(-xrho) - 1)^(-1./xrho)) + ((aF2^(-xrho) + bF2^(-
xrho) - 1)^(-1./xrho));

elseif cop1 == 6;
    r = - (((1-aF1)^xrho + (1-bF1)^xrho - (((1-aF1)^xrho).*( (1-
bF1)^xrho))))^(1 ./ xrho) + (((1-aF2)^xrho + (1-bF1)^xrho - (((1-aF2)^xrho).*(
(1-bF1)^xrho))))^(1 ./ xrho)
        + (((1-aF1)^xrho + (1-bF2)^xrho - (((1-aF1)^xrho).*( (1-
bF2)^xrho))))^(1 ./ xrho) - (((1-aF2)^xrho + (1-bF2)^xrho - (((1-aF2)^xrho).*(
(1-bF2)^xrho))))^(1 ./ xrho);
endif;

if r > 0;
    rnew= ln(r);
else;
    rnew= ln(r-((r.<=0).*(r-.0001)));
endif;

z = rnew;

retp(z);
endp;

```

APPENDIX B: PYTHON CODE FOR CALCULATION OF PSEUDO-ELASTICITIES

```
import numpy as np
import pandas as pd
from scipy.stats import norm,multivariate_normal,mvn
from scipy import integrate
from statsmodels.sandbox.distributions.extras import mvnrmcdf
import os
from numpy.linalg import cholesky

def
calculate_expected_frequencies(individual,data,coef_means,num_thresh_A,num_thr
esh_B,random_draws,number_draws):
    freqs_A = []
    freqs_B = []
    freqs_AB = []
    for this_draw in range(number_draws):
        random_coefs = random_draws[individual][this_draw]
        thresholds_A = random_coefs[:num_thresh_A]
        thresholds_B
        random_coefs[num_thresh_A:num_thresh_A+num_thresh_B]
        propensity_A,propensity_B
    calculate_propensities(coef_means,data,individual)
        corr_term = random_coefs[-1]
        probs_A = calculate_probabilities(propensity_A, thresholds_A)
        probs_B = calculate_probabilities(propensity_B, thresholds_B)
        #probs_AB
    calculate_bivariate_probabilities(propensity_A,propensity_B,thresholds_A,thres
holds_B,corr_term)
        #probs_AB = np.zeros(25).reshape((5,5))+1.0/25
        #if probs_AB.sum() != probs_AB.sum() or (probs_AB.sum() - 1)**2
>= 0.001:
            #this_draw = this_draw - 1
            #print "Probabilities don't add to 1!!!"
            expected_freq_A = (category_numeric_values_A * probs_A).sum()
            expected_freq_B = (category_numeric_values_B * probs_B).sum()
            #expected_freq_AB = (probs_AB * category_numeric_values_AB).sum()
            expected_freq_AB = expected_freq_A + expected_freq_B
            freqs_A.append(expected_freq_A)
            freqs_B.append(expected_freq_B)
            freqs_AB.append(expected_freq_AB)
    freqs_A = np.array(freqs_A)
    freqs_B = np.array(freqs_B)
    freqs_AB = np.array(freqs_AB)
    return (freqs_A,freqs_B,freqs_AB)
```

The Python code presented here was developed by Felipe Ferreira Dias.

```

def
calculate_expected_frequencies_poor(individual,data,coef_means,mean_vals,cov_v
als,num_thresh_A,num_thresh_B,number_draws):
    category_counter_A = np.zeros(categories_A)
    category_counter_B = np.zeros(categories_B)
    category_counter_AB = np.zeros([categories_A,categories_B])
    freq_A = 0.0
    freq_B = 0.0
    freq_AB = 0.0
    for i in range(number_draws):
        random_coefs_are_valid = False
        while random_coefs_are_valid == False:
            random_coefs
np.random.multivariate_normal(mean_vals,cov_vals)
            random_coefs_are_valid
check_draw(random_coefs,num_thresh_A, num_thresh_B)
            thresholds_A = random_coefs[:num_thresh_A]
            thresholds_B
random_coefs[num_thresh_A:num_thresh_A+num_thresh_B]
            propensity_A,propensity_B
calculate_propensities(coef_means,data,individual)
            corr_term = random_coefs[-1]
            random_errors
np.random.multivariate_normal([0,0],[[1,corr_term],[corr_term,1]])
            category_A = classify_category(propensity_A+random_errors[0],
thresholds_A)
            category_B = classify_category(propensity_B+random_errors[1],
thresholds_B)
            category_counter_A[category_A]=category_counter_A[category_A]+1
            category_counter_B[category_B]=category_counter_B[category_B]+1

            category_counter_AB[category_A][category_B]=category_counter_AB[category
_A][category_B]+1
            expected_freq_A = (category_numeric_values_A
category_counter_A.sum()/(1.0*category_counter_A.sum()))
            expected_freq_B = (category_numeric_values_B
category_counter_B.sum()/(1.0*category_counter_B.sum()))
            expected_freq_AB = (category_numeric_values_AB
category_counter_AB).sum()/(1.0*category_counter_AB.sum()))
            freq_A = freq_A + expected_freq_A/(1.0*number_draws)
            freq_B = freq_B + expected_freq_B/(1.0*number_draws)
            freq_AB = freq_AB + expected_freq_AB/(1.0*number_draws)
    return (freq_A,freq_B,freq_AB)

def random_coef_draws(mean_vals,cov_vals,num_thresh_A,num_thresh_B):
    random_coefs_are_valid = False
    while random_coefs_are_valid == False:
        #random_coefs = np.random.multivariate_normal(mean_vals,cov_vals)
        random_coefs = multivariate_normal.rvs(mean_vals,cov_vals)

```



```

        #Cholesky approach
        #pre_draw
        multivariate_normal.rvs(np.zeros(len(mean_vals)),np.identity(len(mean_vals))) =
        #random_coefs = mean_vals + np.dot(cholesky(cov_vals).T,pre_draw)
        #random_coefs[-1]=0.4009
        random_coefs_are_valid = check_draw(random_coefs,num_thresh_A,
num_thresh_B)
        return random_coefs

def check_draw(random_coefs,num_thresh_A, num_thresh_B):
    thresholds_A = random_coefs[:num_thresh_A]
    #print thresholds_A
    thresholds_B = random_coefs[num_thresh_A:num_thresh_A+num_thresh_B]
    #print thresholds_B
    is_valid = True
    for i,left_thresh in enumerate(thresholds_A[:-1]):
        right_thresh = thresholds_A[i+1]
        if left_thresh >= right_thresh:
            is_valid = False
            return is_valid
    for i,left_thresh in enumerate(thresholds_B[:-1]):
        right_thresh = thresholds_B[i+1]
        if left_thresh >= right_thresh:
            is_valid = False
            return is_valid
    #if random_coefs[-1] < -1 or random_coefs[-1] > 1:
    #is_valid = False
    return is_valid

def
populate_random_draws_array(number_obs,number_draws,mean_vals,cov_vals,num_thr
esh_A,num_thresh_B):
    random_draws = []
    for this_obs in range(number_obs):
        random_draws.append([])
        if this_obs %100==0:
            print "Creating random draws for observation number " +
str(this_obs)
            None
            for this_draw in range(number_draws):
                drawn_coefs
                random_coef_draws(mean_vals,cov_vals,num_thresh_A,num_thresh_B) =
                drawn_coefs[-1]=0.4009
                random_draws[this_obs].append(drawn_coefs)
    return random_draws

def
populate_random_draws_array_same_draws_for_all(number_obs,number_draws,mean_va
ls,cov_vals,num_thresh_A,num_thresh_B):
    random_draws = []

```

```

    for this_obs in range(number_obs):
        random_draws.append([])
        if this_obs %100==0:
            print "Creating random draws for observation number " +
str(this_obs)
            None
            for this_draw in range(number_draws):
                if this_obs == 0:
                    drawn_coefs
random_coef_draws(mean_vals,cov_vals,num_thresh_A,num_thresh_B)
                    random_draws[this_obs].append(drawn_coefs)
                else:
                    random_draws[this_obs].append(drawn_coefs)
            return random_draws

def calculate_propensities(coef_means,data,individual):
    this_data
data[individual:individual+1].reset_index().drop(labels=["index"],axis=1)
    all_columns = np.copy(coef_means.Var.values[8:-1])
    all_columns_no_prefix = np.copy(all_columns)
    for i,this_col in enumerate(all_columns):
        all_columns_no_prefix[i] = this_col[2:]
    all_coef_vals = np.copy(coef_means.Mean.values[8:-1])
    num_coefs = len(all_columns)

    filtered_A = np.zeros(num_coefs).astype("bool")
    filtered_B = np.zeros(num_coefs).astype("bool")
    for i in range(num_coefs):
        if "A_" in all_columns[i]:
            filtered_A[i] = True
        if "B_" in all_columns[i]:
            filtered_B[i] = True
    propensity_A = 0
    propensity_B = 0
    cols_A = all_columns_no_prefix[filtered_A]
    coefs_A = all_coef_vals[filtered_A]
    data_A = this_data[cols_A].values[0]
    cols_B = all_columns_no_prefix[filtered_B]
    coefs_B = all_coef_vals[filtered_B]
    data_B = this_data[cols_B].values[0]
    propensity_A = (data_A * coefs_A).sum()
    propensity_B = (data_B * coefs_B).sum()
    return (propensity_A,propensity_B)

def classify_category(propensity, thresholds):
    if propensity < thresholds[0]:
        category = 0
    elif propensity >= thresholds[len(thresholds)-1]:
        category = len(thresholds)
    else:
        for i,left_thresh in enumerate(thresholds[:-1]):

```

```

        right_thresh = thresholds[i+1]
        if propensity >= left_thresh and propensity <
right_thresh:
            category = i+1
            break

    return category

def calculate_probabilities(propensity, thresholds):
    thresh_neg_inf = -300
    thresh_pos_inf = +300
    thresholds = np.array([thresh_neg_inf] + list(thresholds) +
[thresh_pos_inf])
    #print thresholds
    probs = np.zeros(len(thresholds)-1)
    total_prob = 0.0
    for i, left_thresh in enumerate(thresholds[:-1]):
        right_thresh = thresholds[i+1]
        this_prob = norm.cdf(right_thresh-propensity)-
norm.cdf(left_thresh-propensity)
        probs[i] = this_prob
        total_prob = total_prob + this_prob
    #print total_prob
    return probs

def
calculate_bivariate_probabilities(propensity_A, propensity_B, thresholds_A, thresh
holds_B, corr_term):
    thresh_neg_inf = -100
    thresh_pos_inf = +100
    thresholds_A = np.copy(np.array([thresh_neg_inf] + list(thresholds_A) +
[thresh_pos_inf]))
    thresholds_B = np.copy(np.array([thresh_neg_inf] + list(thresholds_B) +
[thresh_pos_inf]))
    probs = np.zeros([len(thresholds_A)-1, len(thresholds_B)-1])
    total_prob = 0
    for k in range(len(thresholds_A)-1):
        left_thresh_A = thresholds_A[k]
        right_thresh_A = thresholds_A[k+1]
        for j in range(len(thresholds_B)-1):
            left_thresh_B = thresholds_B[j]
            right_thresh_B = thresholds_B[j+1]
            b_left = left_thresh_A - propensity_A
            b_right = right_thresh_A - propensity_A
            d_left = left_thresh_B - propensity_B
            d_right = right_thresh_B - propensity_B
            #print b_left, b_right, d_left, d_right
            p
            =
calc_prob_thresh(b_left, b_right, d_left, d_right, corr_term)
            total_prob += p
            #print "k = " + str(k) + " and j = " + str(j) + "      p =
" + str(p)

```

```

        probs[k][j]=p
    #print total_prob
    return probs

def prepare_for_doubleint(x,y,means,cov_matrix):
    return multivariate_normal.pdf([x,y],means,cov_matrix)

def bivariate_normal_cdf(upper_x,upper_y,means,cov_matrix):
    return integrate.dblquad(prepare_for_doubleint,-np.inf,upper_x,lambda x:
-np.inf, lambda x: upper_y,(means,cov_matrix))[0]

def odd_bivariate_normal_cdf(upper_x,upper_y,means,cov_matrix):
    low = np.array([-500, -500])
    upp = np.array([upper_x, upper_y])
    mu = np.array(means)
    S = np.array(cov_matrix)
    p,i = mvn.mvnun(low,upp,mu,S)
    return p

def calc_prob_thresh(b_left,b_right,d_left,d_right,corr_term):
    cov_matrix = [[1,corr_term],[corr_term,1]]
    means = [0,0]
    #return odd_bivariate_normal_cdf(b_right,d_right,means,cov_matrix) -
odd_bivariate_normal_cdf(b_left,d_right,means,cov_matrix) -
odd_bivariate_normal_cdf(b_right,d_left,means,cov_matrix) +
odd_bivariate_normal_cdf(b_left,d_left,means,cov_matrix)
    return mvnrmcdf([b_right,d_right],means,cov_matrix) -
mvnrmcdf([b_left,d_right],means,cov_matrix) -
mvnrmcdf([b_right,d_left],means,cov_matrix) +
mvnrmcdf([b_left,d_left],means,cov_matrix)

def
calculate_expected_freqs_for_all_dataset(data,number_obs,coef_means,mean_vals,
cov_vals,num_thresh_A, num_thresh_B,random_draws,number_draws):
    number_draws = len(random_draws[0])
    freq_A_list = []
    freq_B_list = []
    freq_AB_list = []
    for i in range(number_obs):
        if i%100==0:
            print "Calculating probabilities for observation number "
+ str(i)
            freqs
=
calculate_expected_frequencies(i,data,coef_means,num_thresh_A,
num_thresh_B,random_draws,number_draws)
            #freqs
=
calculate_expected_frequencies_poor(i,data,coef_means,mean_vals,cov_vals,num_t
hresh_A,num_thresh_B,number_draws)
            freq_A_list.append(freqs[0])

```

```

        freq_B_list.append(freqs[1])
        freq_AB_list.append(freqs[2])
    freq_A_list = np.array(freq_A_list)
    freq_B_list = np.array(freq_B_list)
    freq_AB_list = np.array(freq_AB_list)
    results_in_list = np.array([freq_A_list, freq_B_list, freq_AB_list])
    return results_in_list

def get_databases(studied_var, data):
    base_case = 0
    data_sets = []
    for i in range(6):
        data_sets.append(data.copy())

    if studied_var == "driver":
        total_data_sets = 2
        base_case = 0

        #Case 1: Doesn't have a driver's license
        data_sets[0]["LICEN_2"]=1

        #Case 1: Has a driver's license
        data_sets[1]["LICEN_2"]=0
        descriptions = ["Doesn't have a driver's license", "Has a driver's
license"]

    elif studied_var == "employment":
        total_data_sets = 2
        base_case = 1

        #Case 1: Employed (Full or self-employed)
        data_sets[0]["EMPNT_13"]=1

        #Case 2: Unemployed and part-time
        data_sets[1]["EMPNT_13"]=0
        descriptions = ["Employed (Full or self-employed)", "Unemployed
and part-time"]

    elif studied_var == "education":
        total_data_sets = 2
        base_case = 1

        #Case 1: Bachelor's degree or higher
        data_sets[0]["EDUC_6_7"]=1

        #Case 2: Some college or less
        data_sets[1]["EDUC_6_7"]=0
        descriptions = ["Bachelor's degree or higher", "Some college or
less"]

    elif studied_var == "age":

```

```

total_data_sets = 3
base_case = 0

#Case 1: 18-34
data_sets[0]["AGE_7_8"]=0
data_sets[0]["AGE_9P"]=0

#Case 2: 35-54
data_sets[1]["AGE_7_8"]=1
data_sets[1]["AGE_9P"]=0

#Case 3: 55 and above
data_sets[2]["AGE_7_8"]=0
data_sets[2]["AGE_9P"]=1
descriptions = ["18-34","35-54","55 and above"]

elif studied_var == "smartphone_vs_hhstruc":
    total_data_sets = 3
    base_case = 1

    #Case 1: (Has a Smartphone & Multi person HH) & (Has a Smartphone
    & Single person HH)
    data_sets[0]["SPHN_2_3"]=0
    data_sets[0]["SPSING"]=0

    #Case 2: Doesn't have a Smartphone & Multi person HH
    data_sets[1]["SPHN_2_3"]=1
    data_sets[1]["SPSING"]=0

    #Case 3: Doesn't have a Smartphone & Single person HH
    data_sets[2]["SPHN_2_3"]=1
    data_sets[2]["SPSING"]=1
    descriptions = ["(Has a Smartphone & Multi person HH) & (Has a
    Smartphone & Single person HH)","Doesn't have a Smartphone & Multi person
    HH","Doesn't have a Smartphone & Single person HH"]

elif studied_var == "income_vs_kids":
    total_data_sets = 4
    base_case = 0

    #Case 1: (Above $100,000 & No Kids) & ($50,000-$99,999 & No
    kids) & (Above $100,000 & Has kids in HH)
    data_sets[0]["INC12"]=0
    data_sets[0]["KDIN12"]=0
    data_sets[0]["KDIN34"]=0
    data_sets[0]["KDIN14"]=0
    data_sets[0]["NKIN12"]=0

    #Case 2: Below $49,999 & No kids
    data_sets[1]["INC12"]=1
    data_sets[1]["KDIN12"]=0

```

```

data_sets[1]["KDIN34"]=0
data_sets[1]["KDIN14"]=0
data_sets[1]["NKIN12"]=1

#Case 3: Below $49,999 & Has kids
data_sets[2]["INC12"]=1
data_sets[2]["KDIN12"]=1
data_sets[2]["KDIN34"]=0
data_sets[2]["KDIN14"]=1
data_sets[2]["NKIN12"]=0

#Case 4: $50,000-$99,999 & Has kids in HH
data_sets[3]["INC12"]=0
data_sets[3]["KDIN12"]=0
data_sets[3]["KDIN34"]=1
data_sets[3]["KDIN14"]=1
data_sets[3]["NKIN12"]=0
descriptions = ["(Above $100,000 & No Kids) & ($50,000-$99,999 &
No kids) & (Above $100,000 & Has kids in HH)", "Below $49,999 & No
kids", "Below $49,999 & Has kids", "$50,000-$99,999 & Has kids in HH"]

elif studied_var == "resdens_vs_numveh":
    total_data_sets = 5
    base_case = 0

#case 1: (0 Vehicles & Non-dense region) & (0 vehicles & Dense)
data_sets[0]["VEH1"]=0
data_sets[0]["VEH2P"]=0
data_sets[0]["V1NDEN"]=0
data_sets[0]["V2NDEN"]=0
data_sets[0]["RDVEH1"]=0
data_sets[0]["RDVEH2"]=0

#Case 2: 1 Vehicle & Non-dense
data_sets[1]["VEH1"]=1
data_sets[1]["VEH2P"]=0
data_sets[1]["V1NDEN"]=1
data_sets[1]["V2NDEN"]=0
data_sets[1]["RDVEH1"]=0
data_sets[1]["RDVEH2"]=0

#Case 3: 1 Vehicle & Dense
data_sets[2]["VEH1"]=1
data_sets[2]["VEH2P"]=0
data_sets[2]["V1NDEN"]=0
data_sets[2]["V2NDEN"]=0
data_sets[2]["RDVEH1"]=1
data_sets[2]["RDVEH2"]=0

#Case 4: 2 Vehicles & Non-dense
data_sets[3]["VEH1"]=0

```

```

        data_sets[3]["VEH2P"]=1
        data_sets[3]["V1NDEN"]=0
        data_sets[3]["V2NDEN"]=1
        data_sets[3]["RDVEH1"]=0
        data_sets[3]["RDVEH2"]=0

        #Case 5: 2 Vehicles & Dense
        data_sets[4]["VEH1"]=0
        data_sets[4]["VEH2P"]=1
        data_sets[4]["V1NDEN"]=0
        data_sets[4]["V2NDEN"]=0
        data_sets[4]["RDVEH1"]=0
        data_sets[4]["RDVEH2"]=1
        descriptions = ["(0 Vehicles & Non-dense region) & (0 vehicles &
Dense)","1 Vehicle & Non-dense","1 Vehicle & Dense","2 Vehicles & Non-
dense","2 Vehicles & Dense"]
        return data_sets, total_data_sets, descriptions, base_case

#os.chdir("C:\\Users\\ff3883\\Desktop\\elas")
#os.chdir("D:\\temp\\elast")
os.chdir("D:\\temp")

categories_A = 5
categories_B = 5

category_numeric_values_A = [0,0.333,2,4,16]
category_numeric_values_B = [0,0.333,2,4,16]
category_numeric_values_AB =
np.array(category_numeric_values_A).reshape(len(category_numeric_values_A),1)+
np.array(category_numeric_values_B)

num_thresh_A = categories_A - 1
num_thresh_B = categories_B - 1

cov = pd.read_csv("cov_mat.csv")
cov_vals = cov.drop(labels=["Var"],axis=1).values

coef_means = pd.read_csv("mean.csv")
mean_vals = coef_means.Mean.values
for i,this_col in enumerate(coef_means.Var.values):
    coef_means.Var.values[i]=this_col.upper()
usable_columns = np.copy(coef_means.Var.values[8:-1])
for i,this_col in enumerate(usable_columns):
    usable_columns[i]=this_col[2:]

data = pd.read_csv("Puget_Data.csv")
for this_col in data.columns:
    if this_col.upper() not in usable_columns:

```



```

        data = data.drop(labels=[this_col],axis=1)
    else:
        data = data.rename(columns={this_col:this_col.upper()})
number_obs = len(data)

number_draws = 50
random_draws =
populate_random_draws_array(number_obs,number_draws,mean_vals,cov_vals,num_thr
esh_A,num_thresh_B)

#random_draws =
populate_random_draws_array_same_draws_for_all(number_obs,number_draws,mean_va
ls,cov_vals,num_thresh_A,num_thresh_B)
#random_draws_fake = np.zeros([number_obs,number_draws])

possible_vars =
["driver","employment","education","age","smartphone_vs_hhstruc","income_vs_ki
ds","resdens_vs_numveh"]
#
# 0 1 2 3 4
5 6

result_dataframe = pd.DataFrame()
result_names =
["freq_A_elast","freq_B_elast","freq_AB_elast","freq_A_elast_stderr","freq_B_e
last_stderr","freq_AB_elast_stderr","freq_A_diffs","freq_B_diffs","freq_AB_dif
fs","freq_A_diffs_stderr","freq_B_diffs_stderr","freq_AB_diffs_stderr","freq_A
","freq_B","freq_AB","freq_A_stderr","freq_B_stderr","freq_AB_stderr"]
result_dataframe["ResultNames"]=result_names

for k,studied_var in enumerate(possible_vars):
    print "Started analyzing " + studied_var
    data_sets,total_data_sets,descriptions,base_case =
get_databases(studied_var,data)
    raw_results = []
    for i in range(total_data_sets):

        raw_results.append(calculate_expected_freqs_for_all_dataset(data_sets[i]
,number_obs,coef_means,mean_vals,cov_vals,num_thresh_A,
num_thresh_B,random_draws,number_draws))
        raw_results = np.array(raw_results)

        diff_results = []
        percentage_results = []
        for j in range(total_data_sets):
            if j != base_case:
                diff_results.append(raw_results[j]-raw_results[base_case])
                percentage_results.append((1.0*raw_results[j]-
raw_results[base_case])/raw_results[base_case])
            else:

```

```

        diff_results.append(raw_results[base_case]*0.0)
        percentage_results.append(raw_results[base_case]*0.0)
diff_results = np.array(diff_results)
percentage_results = np.array(percentage_results)
final_results = []

for j in range(total_data_sets):
    this_case_freq_A = raw_results[j,0].mean(axis=0).mean()
    this_case_freq_B = raw_results[j,1].mean(axis=0).mean()
    this_case_freq_AB = raw_results[j,2].mean(axis=0).mean()
    this_case_freq_A_stderr = raw_results[j,0].mean(axis=0).std()
    this_case_freq_B_stderr = raw_results[j,1].mean(axis=0).std()
    this_case_freq_AB_stderr = raw_results[j,2].mean(axis=0).std()
    if j != base_case:
        this_case_freq_A_elast =
percentage_results[j,0].mean(axis=0).mean()
        this_case_freq_B_elast =
percentage_results[j,1].mean(axis=0).mean()
        this_case_freq_AB_elast =
percentage_results[j,2].mean(axis=0).mean()
        this_case_freq_A_elast_stderr =
percentage_results[j,0].mean(axis=0).std()
        this_case_freq_B_elast_stderr =
percentage_results[j,1].mean(axis=0).std()
        this_case_freq_AB_elast_stderr =
percentage_results[j,2].mean(axis=0).std()
        this_case_freq_A_diffs =
diff_results[j,0].mean(axis=0).mean()
        this_case_freq_B_diffs =
diff_results[j,1].mean(axis=0).mean()
        this_case_freq_AB_diffs =
diff_results[j,2].mean(axis=0).mean()
        this_case_freq_A_diffs_stderr =
diff_results[j,0].mean(axis=0).std()
        this_case_freq_B_diffs_stderr =
diff_results[j,1].mean(axis=0).std()
        this_case_freq_AB_diffs_stderr =
diff_results[j,2].mean(axis=0).std()
        these_results =
[this_case_freq_A_elast,this_case_freq_B_elast,this_case_freq_AB_elast,this_ca
se_freq_A_elast_stderr,this_case_freq_B_elast_stderr,this_case_freq_AB_elast_s
tderr,this_case_freq_A_diffs,this_case_freq_B_diffs,this_case_freq_AB_diffs,th
is_case_freq_A_diffs_stderr,this_case_freq_B_diffs_stderr,this_case_freq_AB_di
ffs_stderr]

        final_results.append(these_results)
    else:
        these_results = list(np.zeros(12))
        these_results =
[this_case_freq_A,this_case_freq_B,this_case_freq_AB,this_case_freq_A_stderr,t
his_case_freq_B_stderr,this_case_freq_AB_stderr]

```

```

        result_dataframe[descriptions[j]]=these_results
    final_results = np.array(final_results)
    result_dataframe.to_csv("df_test_allvars.csv",index=False)
    #np.savetxt("results_"+studied_var+"_single_draw_.csv",np.array(results)
,delimiter=',')
    #np.savetxt("descriptions_"+studied_var+"_draws_from_r.csv",np.array(des
criptions),delimiter=',',fmt="%s")

result_dataframe.to_csv("df_test_allvars.csv",index=False)

```

REFERENCES

- Baptista, P., S. Melo, and C. Rolim. Energy, Environmental and Mobility Impacts of Car-Sharing Systems. Empirical Results from Lisbon, Portugal. *Procedia-Social and Behavioral Sciences*, Vol. 111, 2014, pp. 28-37.
- Bialik, C., A. Flowers, R. Fischer-Baum, and D. Mehta. *Uber is Serving New York's Outer Boroughs More than Taxis Are*, 2015.
<http://fivethirtyeight.com/features/uber-is-serving-new-yorks-outer-boroughs-more-than-taxis-are/>, Accessed July 10, 2016.
- Buhr, S. *Lyft, Now Worth \$5.5 Billion, Hops into the Autonomous Car Race with General Motors*, 2016. <https://techcrunch.com/2016/01/04/lyft-now-worth-5-5-billion-plans-to-get-into-the-autonomous-car-race-with-general-motors/>, Accessed July 10, 2016.
- Clewlow, R. R., Carsharing and Sustainable Travel Behavior: Results from the San Francisco Bay Area. *Transport Policy*, Vol. 51, 2016, pp. 158-164.
- Coll, M. H., M. H. Vandersmissen, and M. Thériault. Modeling Spatio-Temporal Diffusion of Carsharing Membership in Québec City. *Journal of Transport Geography*, Vol. 38, 2014, pp. 22-37.
- Costain, C., C. Ardron, and K. N. Habib. Synopsis of Users' Behaviour of a Carsharing Program: A Case Study in Toronto. *Transportation Research Part A*, Vol. 46, No. 3, 2012, pp. 421-434.
- Dias, F. F., P. Lavieri, V. M. Garikapati, S. Astroza, R. M. Pendyala, and C. R. Bhat. A Behavioral Choice Model of the Use of Car-Sharing and Ride-Sourcing Services: Online Supplement. Available at:
http://www.cae.utexas.edu/prof/bhat/ABSTRACTS/Carsharing/online_supplement.pdf, 2016.
- Efthymiou, D., C. Antoniou, and P. Waddell. Factors Affecting the Adoption of Vehicle Sharing Systems by Young Drivers. *Transport Policy*, Vol. 29, 2013, pp. 64-73.
- Emmerink, R. H., P. Nijkamp, P. Rietveld, and J. N. Van Ommeren. Variable Message Signs and Radio Traffic Information: An Integrated Empirical Analysis of Drivers' Route Choice Behaviour. *Transportation Research Part A*, Vol. 30, No. 2, 1996, pp. 135-153.
- Firnkorn, J. . Triangulation of Two Methods Measuring the Impacts of a Free-Floating Carsharing System in Germany. *Transportation Research Part A*, Vol. 46, No. 10, 2012, pp. 1654-1672.
- Ferdous, N., N. Eluru, C. R. Bhat, and I. Meloni. A Multivariate Ordered-Response Model System for Adults' Weekday Activity Episode Generation by Activity Purpose and Social Context. *Transportation Research Part B: Methodological*, Vol. 44, No. 8, 2010, pp. 922-943.
- Firnkorn, J., and M. Müller. What Will be the Environmental Effects of New Free-Floating Car-Sharing Systems? The Case of Car2go in Ulm. *Ecological Economics*, Vol. 70, No. 8, 2011, pp. 1519-1528.
- Lyft. *Cities We're In*, 2016. <https://www.lyft.com/cities>, Accessed July 10, 2016.

- Martin, E. W. and S. A. Shaheen. Greenhouse Gas Emission Impacts of Carsharing in North America. *IEEE Transactions Intelligent Transportation Systems*, Vol. 12, No. 4, 2011, pp. 1074-1086.
- Morning Consult. *National Tracking Poll #150505 – June 29–31, 2015*. https://morningconsult.com/wp-content/uploads/2015/06/150505_crosstabs_mc_v2_AD.pdf, Accessed July 12, 2016.
- PSRC. *Puget Sound Regional Travel Study – Report: 2015 Household Travel Survey*, 2015. <http://www.psrc.org/assets/14106/2015-Household-Survey-Tech-Report.pdf>, Accessed July 12, 2016.
- Rayle, L., D. Dai, N. Chan, R. Cervero, and S. Shaheen. Just a Better Taxi? A Survey-Based Comparison of Taxis, Transit, and Ridesourcing Services in San Francisco. *Transport Policy*, Vol. 45, 2016, pp. 168-178.
- SFMTA. *Travel Decisions Survey 2014-Summary Report*, 2014. <https://www.sfmta.com/sites/default/files/reports/2015/Travel%20Decision%20Survey%202014%20Summary%20Report.pdf>, Accessed July 12, 2016.
- Shaheen, S., A. Cohen, and M. Chung. North American Carsharing: 10-year Retrospective. In *Transportation Research Record: Journal of the Transportation Research Board, 2110*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 35-44.
- Shaheen, S. A. Mobility and the Sharing Economy. *Transport Policy*, Vol. 51, 2016, pp. 141-142.
- Smith, A. *Shared, Collaborative and On Demand: The New Digital Economy*, 2016. <http://www.pewinternet.org/2016/05/19/the-new-digital-economy/>, Accessed July 15, 2016.
- Somerville, H. *Uber Reaches 2 Billion Rides Six Months after Hitting its First Billion*, 2016. <http://www.reuters.com/article/us-uber-rides-idUSKCN0ZY1T8>, Accessed July 15, 2016.
- Viechnicki, P., A. Khuperkar, T. D. Fishman, and W. D. Eggers. *Smart Mobility Collection*, 2015. <http://dupress.com/articles/smart-mobility-trends-study-findings/?coll=12925>, Accessed July 18, 2015.
- Waheed, S., L. Herrera, S. Ritoper, J. Mehta, H. Romero, and V. Narro. *Ridesharing or Ridestealing? Changes in Taxi Ridership and Revenue in Los Angeles 2009-2014*, 2015. <http://www.irle.ucla.edu/publications/documents/Taxi-Policy-Brief.pdf>, Accessed July 15, 2016.
- Zoepf, S. M. and D. R. Keith. User Decision-Making and Technology Choices in the US Carsharing Market. *Transport Policy*, Vol. 51, 2016, pp. 150-157.